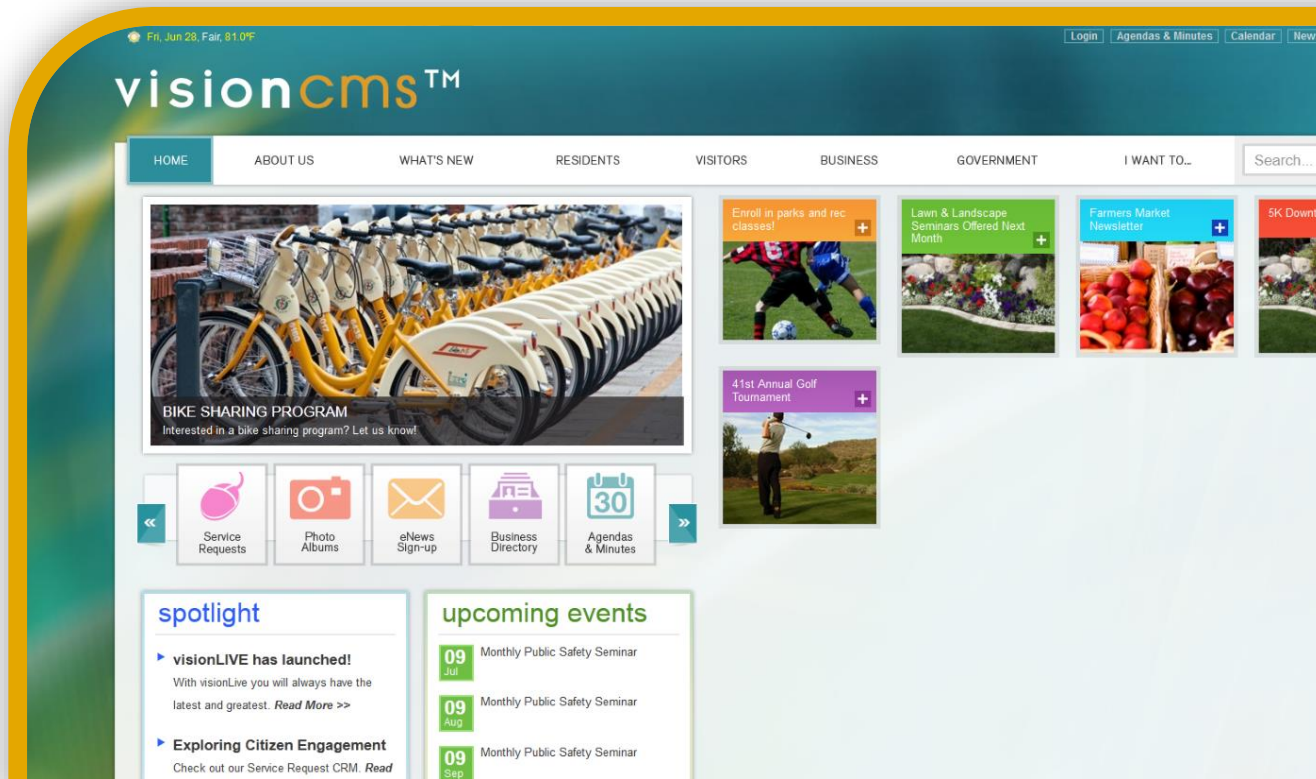


Vision Search Server Configuration



November 3, 2015 – 2015.3.928.0

Table of Contents

Preface.....	2
Important Terminologies.....	2
Search Engines	3
Initialization and Adding a new Search Engine	3
Configuration and Editing a Search Engine	4
Deleting Search Engines	6
Users.....	7
Adding a User.....	7
Editing a User.....	8
Deleting a User.....	8
Security Roles.....	8
Adding a Security Role.....	8
Editing a Security Role.....	8
Deleting Security Roles.....	9
Quick Start Guide to Configuring visionCMS™ for Vision Search.....	9
Enabling Vision Search	9
Adding a Search Engine to visionCMS™	9
Using a Search Engine in a Search Widget in visionCMS™	10



Preface

This document's main purpose is to serve as an instructional material to configure and manage a *Vision Search Server* with *visionCMS™*. This document will not include installation instructions for the *Vision Search Server*.

Important Terminologies

Before starting, it is beneficial to know the terms found in this document. These terminologies are helpful for communication in configuration and troubleshooting.

1. **Crawler / Crawler Service** – *The crawler is in charge of crawling a website and its internal connected links that form a web and send the information to the indexer. A crawler starts from a Root URL and ends crawling until no more links are found, a duplicate has been found in its database or if it achieves the crawler depth setting. Another name for a crawler is “spider”.*
2. **Indexer Service** – *The indexer service or indexer is in charge of reading and analysis of text contained in the page. The indexer is controlled by comment tags `<!--startindex-->` and `<!--stopindex-->` to tell it where to start or stop reading through the page and adding the information to its index. The indexer is capable of reading documents ranging from Microsoft's Word to PDF using intelligent character recognition. The indexer service relies on crawler information before it can start with its task.*
3. **Management Console** – *This is the UI and the heart of the Vision Search Server. This is where all configuration of Search Engines, scheduling and communication between crawler and indexer take place. This site houses the web service that *visionCMS™* communicates to fetch results from the indexer. It is independent of the crawler and indexer and can continue functioning to return results as long as the crawler and indexer have completed their tasks and the management console reports that documents/pages exist in its database.*
4. **Documents** – *This refers to the pages and documents that have been successfully crawled and indexed to the database.*
5. **Regular Expressions** – *Regular expressions are character combinations that are used as a quick algorithm to determine if a string or set of strings match a condition or conditions. Unlike comparing strings, regular expressions provide more flexibility such as what characters or symbols to expect in a sequence. Several tools are available online to create regular expressions with ease.*



Search Engines

The core function of the system, *Search Engines* are collections of pages and documents that contain a plethora of information on a site or several sites.


Initialization and Adding a new Search Engine

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Search**” tab and then click “**Search Engines**”.
3. Click “**Add Search Engine**” to add a *Search Engine*.
4. Fill in the fields below in this page with their specific requirements:
 - **Name** – *The name of the search engine, preferably the domain name for easy identification in the future.*
 - **Description** – *A description of the search engine. This field can also be used to keep notes or other information about the search engine.*
 - **Language** – *Selects which language or dictionary to use when indexing.*
 - **Root URLs** – *A Root URL is the primary entry point of the crawler. This is the first page that the crawler lands and where it will first look for links and start navigating to other pages. Multiple Root URLs can be entered here separating each with a line break by pressing the return key.*
 - **Allow Paths** – *This tells the crawler if it can stay within the current page based off of the specified URL controlled by a regular expression.*
 1. *For example, putting a regular expression “www\.domain\.com\/.*” only allows for URLs that have “www.domain.com” in their URL.*
 2. *It is recommended to be specific when creating regular expressions to prevent crawling of pages that are not included in the index. For example, putting a regular expression “.*” and crawling a site with a link to “www.outdomain.com” in the home page adds the site to the index and its pages in the results. This not only slows down the server but also returns bad search results.*
 - **Disallow Paths** – *This tells the crawler to avoid these URL chunks or paths if found in the URL. This field is opposite of the Allow Paths field as it does not continue crawling on pages that match the regular expression on this field. For example, the regular expression “/Admin\/.*” prevents the crawler to index the backend for security.*



- **Indexing Allow Paths** – *This setting is NOT similar to the Allow Paths. The crawler can still continue crawling all parts of the page but it will not index the page if it does not meet the regular expressions in this field.*
 - **Allow Formats** – *This setting which type of documents the system will return when a search is performed. This is useful for creating special search engines that only perform a set function for example, a set of library books in PDF format but not including HTML or pages.*
5. To continue configuration, the *Search Engine* must be saved for now. To save, click **“Save”** to add the *User* or **“Back”** to cancel.
 6. Proceed to editing the newly add *Search Engine* in the next section.

Configuration and Editing a Search Engine

1. Access the Administration page in the **“Vision Search Management Console”**.
2. Go to the **“Search”** tab and then click **“Search Engines”**.
3. Click  to edit a *Search Engine*.

Note: There will be several tabs visible now with different functions. Some tabs will automatically save after configuration but some have “Save” and “Back” buttons to save and cancel to the previous page.

4. Configure the Settings tab for the *Search Engine*:
 - **Max Document Age** – *This value is in days. It specifies the maximum allowable age of a document in the collection. This setting is useful when allowing only new documents to be indexed. Setting this value to -1 will allow for all pages and documents of any creation or publish date.*
 - **Max Document Size** – *This value is in kilobytes. It specifies the maximum allowable size of a document in the collection. This setting is useful if there are several large documents and there is a significant slowdown in returning results.*
 - **Spider Max. Depth** – *This value is an integer and controls how deep the crawler will go through sub-links in a page. When a page consists of duplicate documents the spider starts checking and counting the depth and stops at the set depth. This prevents an infinite loop between pages that have a link with each other. It is recommended to set this value to 3 in regular configurations.*



- **Spider Delay** – *This value is in milliseconds. It specifies the wait time in milliseconds for the crawler between HTTP requests to a site. This setting helps mitigate the amount of traffic a web server gets. Setting this value to 0 allows the crawler to go immediately to the next link without any hold time.*
- **User Agent** – *This is the user agent name that the crawler runs on.*

Note: Some gateways and firewalls prevent unknown user agents from entering the network for the crawler to properly work. Check and update these settings if it exists in the network.

- **Ignore Robots** – *If the value is set to Yes, the crawler avoids the “robots.txt” file rules and forces a crawl on the link. If set to No, the crawler avoids the pages that match the rules in the file. More information about “robots.txt” is available online.*
- **Follow Redirects** – *If the value is set to Yes, the crawler will continue even if it encounters a 301 or 302 HTTP code. This is common when a page has been moved to another location or another domain. If set to No, the crawler neglects the redirect code.*
- **Remove Duplicates** – *If set to Yes, the Management Console will not return results with the same path or URL when a search occurs. If set to No, duplicates will appear on the search results.*
- **Stemming** – *If set to Yes, word stemming will be enabled. A good example for word stemming is when a search is performed and the search word is singular but the document contains plural words. In this case if word stemming is enabled, searching for “mark” will return results that contain the word “marks” but will not return the word “marker” as it is a noun and not a verb. If set to No, the search results will return unrelated words disregarding its meaning.*
- **Spelling Suggestions** – *If set to Yes, a spell index is created after the index process. This allows for search term suggestions returned in the search page. If set to No, these search term suggestions will not appear.*


5. Configure the Advanced tab for the Search Engine:

- **Search Engine Key** – *One of the two part authentication hash needed to verify and get the correct Search Engine from the list to get search results.*
- **Search Engine Secret** – *The second part of the two part authentication hash.*
- **Content Type Rule** – *This regular expression for a URL path is used to create custom categories within the search page. All pages and documents matching the URL regular expression will fall under the category name chosen.*



- **Content Type** – *This determines the content types available on the visionCMS™ site. This information is communicated from the visionCMS™ to the Vision Search Server. The default Content Types cannot be modified or deleted, however additional new or custom Content Types can be added.*
6. Configure the Search Features tab for the *Search Engine*:
- **Search Synonyms** – *Custom synonyms can be added here. Synonyms can be used to associate words that are not normally categorized as synonyms but bear significance to the search requirements. For example, regular synonyms “tall” and “high”. Required synonyms such as “parking ticket” and “fees”.*
7. The next step is to create a schedule and or manually force a process to start the crawling and indexing process. This is under the Indexing tab for the *Search Engine*:
- **Index** – *This is the primary function or operation to start indexing the set configuration for the Search Engine. This will start the crawling and then indexing phase. If executed consecutively, existing documents will be deleted and re-indexed.*
 - **Refresh** – *This operation will update the index for documents that have been updated since the time it was last indexed. If a deleted document has been crawled, these are removed from the index. This by definition is the update function for the index.*
 - **Optimize** – *Optimize improves the current index for fast searching. This function is best carried out as a weekly operation at the end of an Index or Refresh operation to improve performance.*
 - **Clear** – *This operation removes and purges the current documents in the Search Engine index. If updating the index, it is best to use Refresh instead of Clear and then index.*
8. Under the Schedule Activities section, a manual operation can be scheduled immediately. If this is a new *Search Engine* it is recommended to perform an *Index* and a *Clear* operation to gauge the performance of the crawler and index on the site and make adjustments accordingly.

Deleting Search Engines

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Search**” tab and then click “**Search Engines**”.
3. Choose a *Search Engine* and click  to delete.



4. Alternatively, to delete multiple *Search Engines* at once, activate the checkbox to the left of each entry and the “**Delete Selected**” button will be enabled. Click this button and it will display a confirmation message that upon clicking “**OK**” will delete the selected entries.

Users


Users allow for history and change tracking in the system. *Users* can be assigned as *Super Users* to gain full control of the *Vision Search Server* and all its features. On the other hand *Users* can be assigned to *Security Roles* and have a specific task or specific tasks within the system.

Adding a User

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Admin**” tab and then click “**Users**”.
3. Click “**Add User**” to add a *User*.
4. Fill in the fields below in the “**User Information**” tab with their specific requirements:
 - **First Name** – *The first name of the user.*
 - **Last Name** – *The last name of the user.*
 - **Email** – *The email address of the user. Must be in a valid email address format e.g. <account>@<host>.*
 - **User Name** – *The user name of the user. The user name must be alphanumeric and cannot contain spaces.*
 - **Password** – *The password of the user. Must contain at least one capitalized letter, a numeral and a symbol.*
 - **Confirm Password** – *A repetition of the password for validation before saving.*
 - **Active** – *A flag that determines if the user can still login and be granted access.*
 - **Super User** – *A flag that determines if the user can access the entire system without restriction.*
5. Optional: If the *User* is not a *Super User* a security role can be chosen in the “**Assign User To Security Role**” tab.
6. Click “**Save**” to add the *User* or “**Back**” to cancel.



Editing a User

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Admin**” tab and then click “**Users**”.
3. Click  to edit a *User* and perform the necessary changes.
4. Click “**Save**” to modify the *User* or “**Back**” to cancel.

Note: Information about the fields located in this page can be located under “Adding a User” in the previous section.

Deleting a User

It is not possible to delete a *User* in Vision Search due to system constraints and history logging. To prevent this edit a *User* and turn off the *Active* flag to disable system access.

Security Roles

Security Roles allows for precise control of access in the system. *Security Roles* can control settings, configuration, adding, editing and deleting on a per *User* basis.

Adding a Security Role

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Admin**” tab and then click “**Security Roles**”.
3. Click “**Add Security Role**” to add a *Security Role*.
4. Choose which actions the *User* can perform by activating the checkbox beside it in the “**System Level Permissions**” tab.
5. In the “**Assign Security Role to Users**” tab *Users* can be selected to be assigned to this *Security Role*.
6. Click “**Save**” to add the *Security Role* or “**Back**” to cancel.


Editing a Security Role

1. Access the Administration page in the “**Vision Search Management Console**”.
2. Go to the “**Admin**” tab and then click “**Security Roles**”.



3. Click **“Add Security Role”** to edit an existing *Security Role* and make the necessary modifications.
4. Click **“Save”** to modify the *Security Role* or **“Back”** to cancel.

Deleting Security Roles

5. Access the Administration page in the **“Vision Search Management Console”**.
6. Go to the **“Admin”** tab and then click **“Security Roles”**.
7. Choose a *Security Role* and click  to delete.
8. Alternatively, to delete multiple *Security Roles* at once, activate the checkbox to the left of each entry and the **“Delete Selected”** button will be enabled. Click this button and it will display a confirmation message that upon clicking **“OK”** will delete the selected entries.

Quick Start Guide to Configuring visionCMS™ for Vision Search

After configuring a *Search Engine* and running the crawler and indexer process with documents results, the next step is to add a *Search Engine* to the *visionCMS™* backend. This guide will primarily focus on getting the newly created *Search Engine* to be accessible in *visionCMS™* and be used in a *Search Widget*. In depth information on these is available in the *visionCMS™ Online Reference*.

Enabling Vision Search

1. Access the Administration page in *visionCMS™*.
2. Go to the **“Development”** tab and then click **“System Settings”**.
3. Turn on **“Vision Search”** under the list of *System Settings*.

Note: Only developers have access to this tab.

Adding a Search Engine to visionCMS™

4. Access the Administration page in *visionCMS™*.
5. Go to the **“Site Settings”** tab then under that navigate to **“Site Configurations”** and then click **“Search Management”**.
6. Click **“Add Vision Search”** to add a *Vision Search – Search Engine*.



7. Fill in the fields below in the **“User Information”** tab with their specific requirements:
 - **Name** – *A friendly name given to the search engine just as an identifier.*
 - **Server URL** – *The URL or domain of the Vision Search web server and website. Web services will use this URL to send queries to the Search Server.*
 - **Search Engine Key** – *The Search Engine Key of the Search Engine in the Vision Search Server.*
 - **Search Engine Value** – *The Search Engine Secret of the Search Engine in the Vision Search Server.*
8. Click the **“Test Connection”** button to verify that the configuration is correct and that the visionCMS™ site is able to send and receive queries from the Vision Search Server.
9. Click **“Save”** to add the *Vision Search – Search Engine* or **“Back”** to cancel.

Using a Search Engine in a Search Widget in visionCMS™

1. Create a page or look for an existing page in *visionCMS™* with a *Search Widget* in the *Page Template*.
2. Go to the **“Settings”** of the widget by clicking the wrench icon and under **“Select View”** choose **“Vision Search”**.
3. Choose the newly created *Search Engine* under **“Search Engine”** in the widget settings.
4. Make the necessary configurations to the widget and click **“Save”**.

